

Water Resources Research

RESEARCH ARTICLE

10.1029/2020WR027227

Key Points:

- We propose a data-driven approach to model spatial-temporal variations of water quality
- The approach combines data-dimension reduction method and deep-learning techniques
- The application for DO in Chesapeake Bay shows high model performance

Supporting Information:

- Supporting Information S1

Correspondence to:

X. Yu,
xinyu@vims.edu

Citation:

Yu, X., Shen, J., & Du, J. (2020). A machine-learning-based model for water quality in coastal waters, taking dissolved oxygen and hypoxia in Chesapeake Bay as an example. *Water Resources Research*, 56, e2020WR027227. <https://doi.org/10.1029/2020WR027227>

Received 30 JAN 2020

Accepted 10 AUG 2020

Accepted article online 25 AUG 2020

A Machine-Learning-Based Model for Water Quality in Coastal Waters, Taking Dissolved Oxygen and Hypoxia in Chesapeake Bay as an Example

Xin Yu¹ , Jian Shen¹ , and Jiabi Du² 

¹Virginia Institute of Marine Sciences, College of William and Mary, Gloucester, VA, USA, ²Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, MA, USA

Abstract Hypoxia is a big concern in coastal waters as it affects ecosystem health, fishery yield, and marine water resources. Accurately modeling coastal hypoxia is still very challenging even with the most advanced numerical models. A data-driven model for coastal water quality is proposed in this study and is applied to predict the temporal-spatial variations of dissolved oxygen (DO) and hypoxic condition in Chesapeake Bay, the largest estuary in the United States with mean summer hypoxic zone extending about 150 km along its main axis. The proposed model has three major components including empirical orthogonal functions analysis, automatic selection of forcing transformation, and neural network training. It first uses empirical orthogonal functions to extract the principal components, then applies neural network to train models for the temporal variations of principal components, and finally reconstructs the three-dimensional temporal-spatial variations of the DO. Using the first 75% of the 32-year (1985–2016) data set for training, the model shows good performance for the testing period (the remaining 25% data set). Selection of forcings for the first mode points to the dominant role of streamflow in controlling interannual variability of bay-wide DO condition. Different from previous empirical models, the approach is able to simulate three-dimensional variations of water quality variables and it does not use in situ measured water quality variables but only external forcings as model inputs. Even though the approach is used for the hypoxia problem in Chesapeake Bay, the methodology is readily applicable to other coastal systems that are systematically monitored.

1. Introduction

Hypoxia or low dissolved oxygen (DO) condition is one of the most critical environmental problems in coastal waters. DO concentration less than 2 mg/L is typically considered as the threshold for hypoxia, even though there are studies suggesting that the criterion should be region- and organism-specific (e.g., Vaquer-Sunyer & Duarte, 2008). Hypoxic conditions could cause mortality of aquatic organisms, change biogeochemical cycles, alter the ecosystem community, and reduce fishery yield (Diaz & Rosenberg, 2008). Well-known hypoxic zones include the Chesapeake Bay (Kemp et al., 2005), Northern Gulf of Mexico (Bianchi et al., 2010; Rabalais et al., 2002), Baltic Sea (Conley et al., 2002), and oxygen minimum zones in tropical oceans (Karstensen et al., 2008). There have been significant changes in the duration and intensity of hypoxia in estuaries and coastal waters during the past few decades, raising great concerns from environmental management agencies. Since 1950, more than 500 sites in coastal waters have reported hypoxic conditions, and fewer than 10% of these systems were known to have hypoxia before 1950 (Breitburg et al., 2018). The rapid spread of hypoxia worldwide is generally attributed to the increase of nutrient loading in the past decades (Diaz & Rosenberg, 2008). Recent studies suggest climate change (e.g., warming and changing wind field) and the resultant change in physical conditions (e.g., oxygen solubility and estuarine circulation) also contribute to the worsened hypoxic conditions in lakes, estuaries, and coastal waters (Carstensen et al., 2014; Deng et al., 2018; Du et al., 2018; Scully, 2010; Wilson et al., 2015).

Chesapeake Bay, the largest estuary in the United States, was noted to have hypoxic conditions back to the 1930s (Newcombe & Horne, 1938) and has seen an increase of hypoxia over the past century (Hagy et al., 2004) (Figure 1). Climatological condition at an upper bay station (CB3.3C) based on a 32-year record (1985–2016) shows hypoxia starts in late April and ends in middle September (Figure 2) and summer hypoxic area covers about 150 km along the bay's main axis (Figure 1c). The seasonal hypoxia is generally

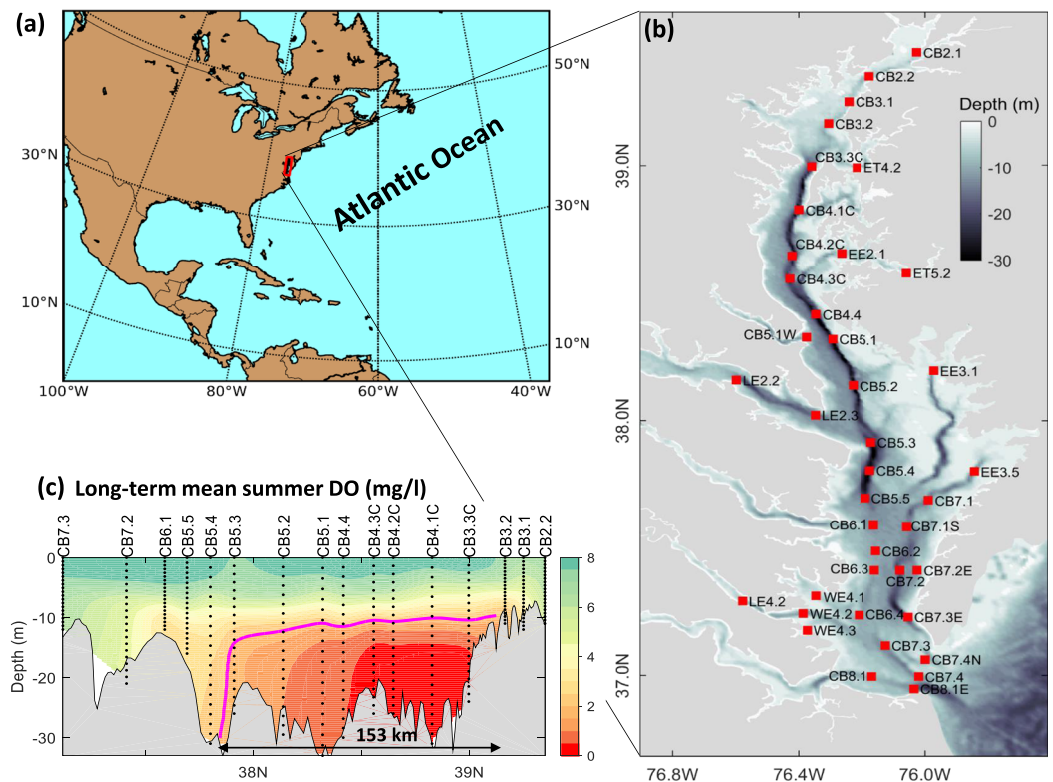


Figure 1. (a) Map of North America, with a red rectangle showing the location of Chesapeake Bay. (b) The 40 long-term (1985–present) Chesapeake Bay Program monitoring stations. Data at these stations have less than 10% data gap and are used in this study. (c) The 32-year mean of the summer (June–August) DO concentration along the main axis of the bay, with the thick magenta line denoting the 2 mg/L (i.e., hypoxia threshold).

believed to be caused by the seasonal growth-settling-decay cycle of phytoplankton and the significant seasonality in water column stratification and air temperature (Taft et al., 1980). Stimulated by the winter-spring pulse of freshwater and nutrient input, algae bloom in spring, resulting in a large amount of organic matter settling down to the bottom water in late spring. The subsequent intense DO consumption during the summer, together with stronger stratification and higher temperature, causes the imbalance of DO supply and consumption in the water column, leading to the depletion of bottom DO (Kemp et al., 2005; Malone et al., 1986; Shen et al., 2013).

Hosting one of the most productive ecosystems on Earth, estuaries and coastal waters have received a lot of attention and are thus the hot spots for environmental researches. Extensive studies have been carried out to understand how DO in coastal waters responds to external forcings with various timescales, such as tide, freshwater discharge, wind, and climate variations (e.g., Fennel & Testa, 2019; Hong & Shen, 2012; Meier et al., 2012; Scully, 2010). Multiple modeling approaches ranging from statistical to fully mechanistic models have been introduced to simulate and predict DO in Chesapeake Bay, and have led to significant advances in understanding the controlling factors for the hypoxia variations. Murphy et al. (2011) used regression models and showed the early summer hypoxic volume in Chesapeake Bay was significantly correlated with stratification strength. Using a cross-wavelet analysis, Muller and Muller (2015) built a neural network model to predict future hypoxic volume of the bay and revealed an antiphase relationship between southwesterly winds and hypoxic volume. With a 3-D numerical model and assuming a constant DO consumption rate, Scully (2010) identified the important control of lateral circulation induced by wind on the bottom DO replenishment. Using a biogeochemistry model, Da et al. (2018) demonstrated the comparable importance of dissolved inorganic nitrogen inputs from the atmosphere and from the adjacent continental shelf. A recent study by Ross and Stock (2019) used a machine learning technique and found the column stratification as the strongest predictor for bottom DO. However, it is still very challenging to accurately simulate or

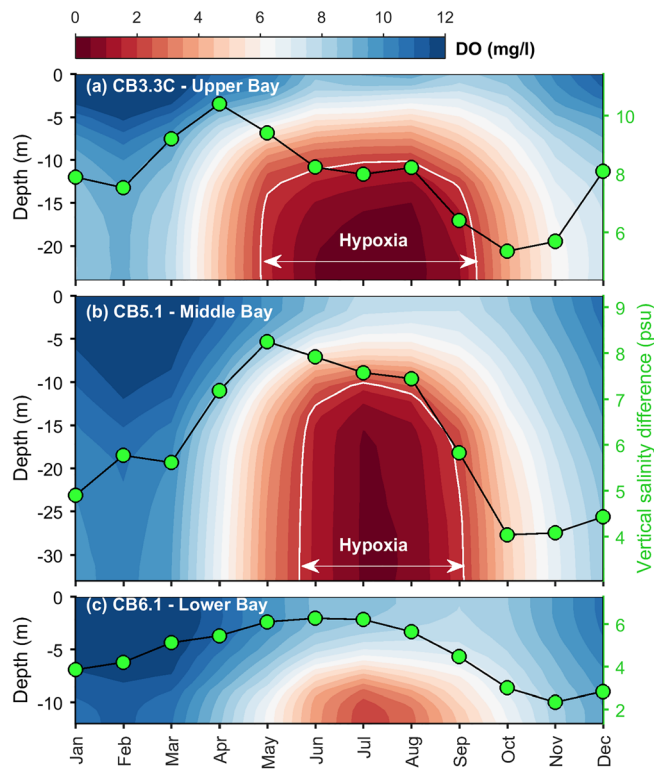


Figure 2. Seasonality of the vertical profile of dissolved oxygen concentration and stratification at selected three mainstem stations, representing the (a) upper bay, (b) middle bay, and (c) lower bay, respectively (see Figure 1 for the location of these stations). White lines in each panel indicate the 2 mg/L contour line, while green dots show the seasonal variation of stratification characterized by the difference between bottom and surface salinity.

predict DO in estuarine and coastal waters due to several reasons: (1) hydrodynamics in estuarine and coastal waters are probably among the most complex dynamical processes on Earth as they are affected by not only natural processes but also anthropogenic activities and they are sensitive to external perturbations such as flooding and storm events; (2) long-term and comprehensive data set are still not readily available for most estuaries and coastal waters even though more field observations have become available recently; a comprehensive data set for coastal water quality covers information of essential forcings (e.g., wind, air temperature, and nutrient fluxes) imposed on the land-water, air-water, and estuary-ocean interfaces of a given coastal water body; (3) low DO problems are caused by complex biochemical processes that vary spatially and temporally. Seeking new approaches to advance the modeling of DO is thus of great interest.

With more available observation data and advances in machine-learning techniques, big-data modeling has been applied in a variety of fields. The new technique provides us an opportunity to improve simulation accuracy and further advance our understanding of environmental problems in estuarine and coastal waters. Data-driven or machine-learning-based modeling have already been applied for storm-surge prediction (Bajo & Umgiesser, 2010), rainfall-runoff processes (Campolo et al., 1999; Hsu et al., 1995), water level and flooding prediction (Campolo et al., 1999; Chang & Chen, 2003), satellite-data retrieval (Keiner & Yan, 1998; Krasnopolsky, 2007; Vilas et al., 2011), marsh classification (Morris et al., 2005), and algal bloom modeling (Muttill & Chau, 2006; Recknagel, 2001; Shen et al., 2019). Neural network application has been applied in Chesapeake Bay as early as 1996 by Scardi (1996) who trained an empirical model for primary phytoplankton production. One of the advantages of the data-driven model is its computa-

tional efficiency. It requires much less computational power compared to complex three-dimensional mechanic models. Therefore, the technique will likely provide us an efficient way for predicting water quality in estuarine and coastal waters.

Here we propose a machine-learning-based data-driven model and examine its performance in simulating the DO condition and hypoxic volume in Chesapeake Bay. Systematical measurement of DO in the bay and continuous monitoring environmental data (e.g., wind, river discharge, air temperature, and nutrient fluxes) since 1984 were publically available. The comprehensive data set in Chesapeake Bay makes the estuary a perfect test site for developing data-driven models. We collected monitoring data of DO from the Chesapeake Bay Program (Figure 1b) and forcing data from a variety of reliable and publically accessible sources to examine the applicability, robustness, and limitations of the data-driven model. The paper is organized as follows. Section 2 describes the method, including the framework of the data-driven model and data collections. Section 3 presents the simulation of hypoxic volume and spatial-temporal variations of DO. The robustness and limitations of the data-driven approach will be discussed in section 4, followed by a short summary.

2. Methods

2.1. Overall Framework of the Proposed Data-Driven Model

The proposed data-driven model includes three major components: empirical orthogonal functions (EOF) analysis, automatic selection of forcing transformation (ASFT), and machine-learning (neural network) (Figure 3). Observed values of the target variable, DO in this study, are first interpolated into a defined vertical grid. The observed target variable, together with forcing variables, will be split into training and testing data sets at the beginning. We used the first 75% for the training and the remaining 25% for the testing

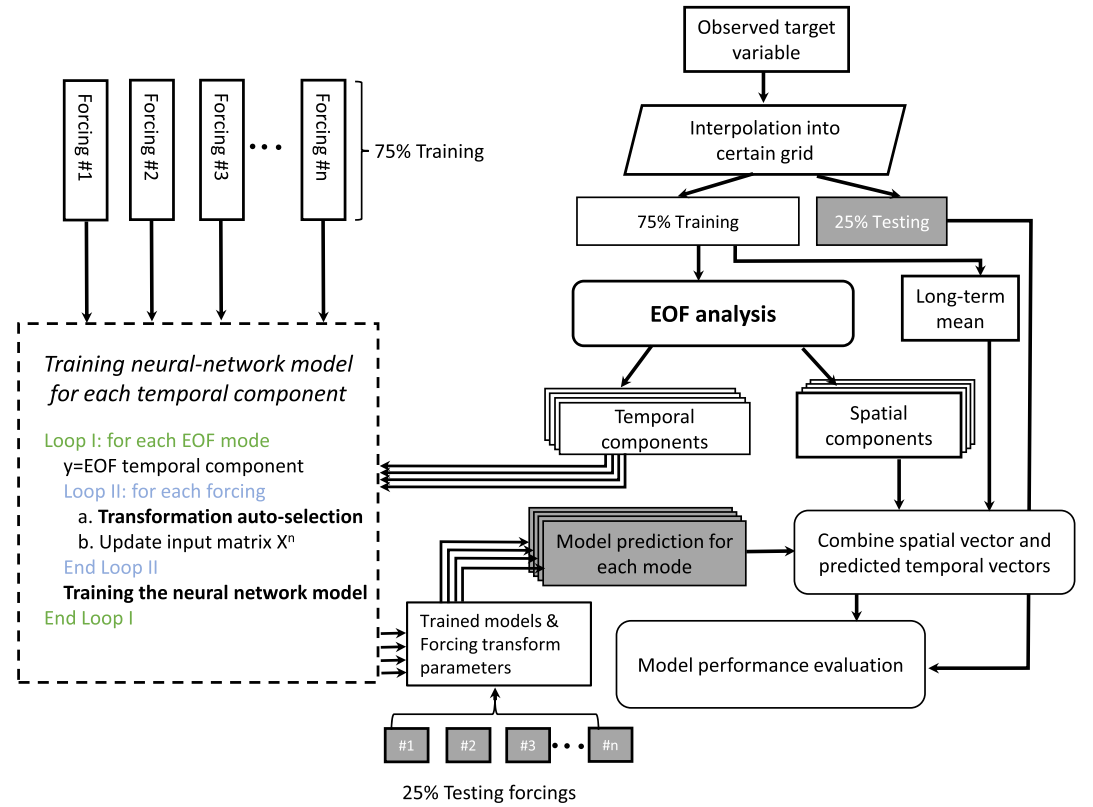


Figure 3. A sketch diagram showing the workflow of the data-driven model.

(Figure 4). In the training data set of the target variable, the long-term mean is extracted and it will be used later to reconstruct the 3-D structure. An EOF analysis is applied to decompose the spatial and temporal components of the target variable in the training data set. For each major principal component extracted from EOF, a neural network model is trained with input forcings selected from the ASFT module. The ASFT is designed to search for the proper transformation of the input forcings. After the model is trained, forcings from the remaining 25% testing data set are transformed based on the transformation function determined by ASFT, and the transformed forcings are put into the trained neural network model to predict the major EOF modes for the testing period. The prediction of the target variable will be obtained as the sum of predicted temporal values multiplying with the corresponding spatial value (also known as the map) and the long-term mean value. The predicted value of target variable, as a function of space and time, will be calculated as

$$C(x, y, z, t) = C_o + \sum_{\text{mod}=1}^N \text{Map}_{\text{mod}}(x, y, z) \times PC_{\text{mod}}(t), \quad (1)$$

where C_o is the long-term mean extracted from the training data set, $\text{Map}(x,y,z)$ and $PC(t)$ are the spatial and temporal values for a given mode, respectively. N is the number of principal components that are trained by the neural network model. In the following sections, the three major components of the proposed data-driven model will be explained in more detail.

2.2. EOF to Reduce Data Dimension

By selecting principal modes based on EOF analysis, the data dimension can be efficiently reduced. The EOF analysis in this study is based on the singular value decomposition algorithm, which decomposes the data matrix F into the following form:

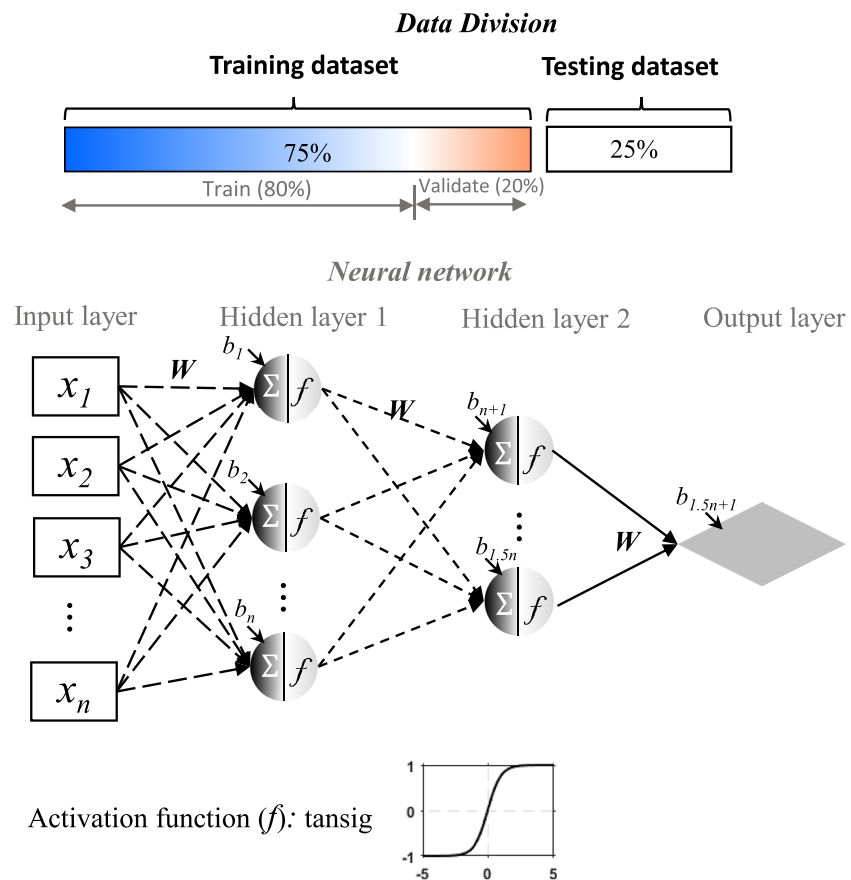


Figure 4. A sketch diagram showing the data-division scheme, the structure of a sample neural network, and the type of activation function used in the proposed data-driven model. For the data-division scheme, the full data are separated into training and testing subdata set; within the training data set, 80% and 20% are used to “train” and “validate” by the neural network.

$$F = U \cdot D \cdot V^T, \quad (2)$$

where U is an orthogonal matrix of temporal vectors, V is an orthogonal matrix of spatial eigenvectors (referred to as maps), and D is a diagonal matrix of the eigenvalues. $U \cdot D$ will give the EOF time series. The data matrix F is arranged with first dimension indicating spatial location and second dimension indicates time (i.e., $F(i,j)$ with i indicate the location and j indicate the time). In this study, the data matrix F has a dimension of 800×288 , representing the spatially varying values at 40 monitoring stations (monthly vertical profiles of DO at each station are interpolated vertically into 20 layers) over the 24 years (1985–2008) in the training data set.

The major variance of the target variable can be well represented by major principal components. For the DO in Chesapeake Bay, our analysis below shows that the first EOF mode accounts for 87% of the total variance and the first five modes account for 93% of the variance. Instead of developing models for each station and each layer (40×20), we choose to fit five regression models for the first five modes to simulate the major variations of the target variable. This strategy will significantly enhance the computational efficiency without losing the major signals.

2.3. Forcing Selection and Transformation

As an essential part of the data-driven model, data of external forcings were carefully collected. Relevant forcings were selected including nutrient loading, river flow, air temperature, solar radiation, and wind speed and direction. River flow and wind are long known to regulate the stratification, estuarine circulation, and water exchange between ocean and estuaries (Hagy et al., 2004; Murphy et al., 2011; Scully, 2010), while

Table 1
A List of the Transformation Options in the Data-Driven Model

Transformation	Subtypes	Formula
Time-lag transformation	1–7	$\phi(t) = x(t\text{-lag})$, with lag ranging within 0, 10, ... 60 days
Accumulative transformation	1–13	$\phi(t) = \text{mean}(x(\tau))$, where $\tau \in [t1 - acc, t2]$, with acc ranging from 0 to 120. $t1$ and $t2$ are the beginning and end of each month; $t1 = t - 15$ and $t2 = t + 15$
Regular transformation	1	$\phi = x$
	2	$\phi = \log(x)$
	3	$\phi = 1/x$
	4	$\phi = \exp((x - \text{mean}(x))/\text{std}(x))$
	5	$\phi = x/(p50 + x)$, also known as Monod-type filter
	6	$\phi = x/(p75 + x)$
	7	$\phi = x/(p25 + x)$
	8	$\phi = (x - \text{mean}(x))/\text{std}(x)$

Note. x = forcing variable; ϕ = transformed forcing variable; t = time; $\text{std}(x)$ = the standard deviation of x ; $\text{mean}(x)$ = the mean value of x ; P25, P50, P75 = the 25, 50, and 75 percentile of x .

nutrient loading and solar radiation are generally regarded as the dominant factors controlling the algal growth. The input forcings are almost the same as required by a three-dimensional ecosystem model (e.g., Cerco & Noel, 2013). Forcing data were collected from various sources, including long-term monitoring programs and reliable reanalysis atmospheric model outputs. River flow and nutrient loadings of the large tributaries, Susquehanna, Potomac, James, and Choptank Rivers, were extracted from USGS (<https://www.usgs.gov/>). As freshwater discharge from other small tributaries are highly correlated with these major rivers, freshwater and nutrients from the small tributaries are not included. Air temperature at Chesapeake Bay Bridge Tunnel station (CBBT, Station ID: 8638901) was extracted from a NOAA database (<https://tidesandcurrents.noaa.gov/>), with data gaps filled with measured values from a nearby station, Cape Henry (Station ID: 8638999). For the wind data, instead of using a continuous monitoring data at a limited number of NOAA stations, we used the global ERA5 reanalyzed wind produced by European Centre for Medium-Range Weather Forecasts (ECMWF: <https://www.ecmwf.int/>), which has a full coverage of the entire Chesapeake Bay with a spatial resolution of 0.25° (total of 33 grid points within the Chesapeake Bay are selected; Figure S1 in the supporting information) and an hourly temporal resolution. Reanalysis wind field from reliable atmospheric models are widely used for 3-D hydrodynamic and water quality models (e.g., Du et al., 2019; Testa et al., 2014; Ye et al., 2018). Taking the bay mouth station CBBT for instance, the observed wind is highly consistent with the ECMWF reanalysis wind (Figure S2).

One feature that makes the model different from previous ones (e.g., Scardi & Harding, 1999; Shen et al., 2019) is that an auto-selection tool for forcing transformation is developed to find the suitable forcing transformation for the model to account for underlying mechanisms with which water quality state variables respond to external forcings. Forcing transformation is a necessary preprocessing step for a data-driven model. Through forcing transformation, input forcings will be converted with the same temporal and spatial resolution, and, more importantly, some particular effects that are ubiquitous in estuarine dynamics, such as time-delay effect and accumulative effect, can be included. For example, it is commonly agreed that summer hypoxia in Chesapeake Bay is attributable to January–May nutrient load instead of summer nutrient load (Murphy et al., 2011). The DO's responses to nutrient load are regulated by not only time-lag effect but also accumulative effect, which is physically meaningful as nutrients from upstream take months to reach middle or lower bay (Shen & Wang, 2007). In the proposed model, we include seven types of time-lag, 13 types of accumulative average, and eight types of transformation functions including log, exponential, Monod-type filter, and normalization (Table 1). In total, there are about 700 combinations of transformation. Furthermore, not all the input forcings are responsible for the DO variations and it is necessary to filter out the unnecessary forcings. The goal of ASFT module is to select the responsible forcings and search for the appropriate transformations for each selected forcing.

In the ASFT module, multiple linear regression is used to find forcings and transformations that can maximize the performance of the model to explain the target variable. Set the target variable as $Y(n \times 1)$ and

the input forcing matrix \mathbf{X} to be empty at the beginning. First, the correlation coefficients (R^2) between \mathbf{Y} and all available forcing variables in all available transformations are computed. The forcing variable with transformation that gives the largest R^2 out of all possible forcings and transformations is selected as the first variable and stored in $\mathbf{X}(:, 1)$. ASFT adds the second forcing variable to \mathbf{X} from the remaining forcings based on the R^2 from multiple linear regressions. Note that during the second round of forcing selection, there will be two forcings in \mathbf{X} , with the first one fixed and the second one chosen from all possible transformed value of the remaining forcings. The new variable will be selected only when the new R^2 is larger than the previous R^2 by at least 0.005. This process will continue until increase of R^2 is less than 0.005. The process is to find a set of independent variables that has the largest contribution to the target variable, while avoiding degrading the model performance when including a large number of variables with high covariance among forcings. The multiple linear regression may not be the best method as it only considers linear relationships, and could be further improved in the future. For the current study, it works well for our problem. With advances in understanding of the underlying mechanisms, additional transformations can be further introduced.

2.4. Neural Network

After selecting the forcings and corresponding transformations, neural network models are trained for the five primary EOF modes (Figure 4). Artificial neural networks are computational models inspired by the functioning of the human brain (Paliwal & Kumar, 2009; Scardi & Harding, 1999). They are composed of a number of “neurons”, the basic computational unit, which takes inputs (\mathbf{x}) from other neurons or external sources, calculates the corresponding weight (\mathbf{w}) for each input, sums the product of weights and input values ($\Sigma \mathbf{w}\mathbf{x}$), plus bias (b), and finally passes this value ($b + \Sigma \mathbf{w}\mathbf{x}$) to an activation function (Figure 4). A number of neurons constitute a hidden neural layer, and a network can have multiple hidden layers. Hyperbolic tangent sigmoid function is used as the activation function in this study, which converts the input value to a value ranging between -1 and 1 . To train a neural network is to get the optimal weights and bias so that the cost function (i.e., the model error) approaches its minimum. Here we used the Levenberg-Marquardt backpropagation training function (Marquardt, 1963), which approaches a second-order training speed without computing the Hessian matrix directly and appears to be the fastest method for training moderate-sized feedforward neural network (Hagan & Menhaj, 1999). The Matlab Neural Network Toolbox (version 10.0) was used for this study. The training process will stop when any of the following conditions occurs: (1) number of epochs reach the defined maximum epochs (set to be 100); (2) cost function (mean square error) is minimized to the 0; (3) performance gradient falls below $1e-7$. For details of the algorithm of Levenberg-Marquardt method, readers are referred to the help document in Matlab. The algorithm is widely recognized and well implemented in the Matlab software.

As mentioned in section 2.1, the overall division of the full data set in the proposed approach is configured as follows: the last 25% of data are reserved for the testing; the remaining 75% are used for the training. The neural network, internally, is also set to randomly divide the input data set (i.e., the training data set) into two independent portions, the training and validation portions, which accounts for 80% and 20% of the input data length (i.e., 60% and 15% of the full data records), respectively. By default, the toolbox randomly chooses the “train” and “validate” portions for each training, resulting in slightly different neural network parameters and thus different predictions for the testing period. To address the related uncertainties, we train the neural network model for 100 times for each principal component, use the ensemble mean of these models as the final prediction, and use the standard deviation of these models’ predictions to quantify the uncertainties.

We use two hidden layers, with a neuron number of N and round up of $N/2$ for the first and second hidden layers, respectively, where N is the number of input forcings. The number of input forcings varies for different principal EOF components and is determined in the forcing selection (see section 2.3). Sensitivity tests (not shown) regarding the number of hidden layers and the number of neurons do not show significantly different results. It is believed the forcing selection is more important than the hidden layer configuration for water quality problems.

2.5. Model Performance Evaluation

Besides the common statistical measures including root-mean-square error (RMSE) and coefficient of determination (R^2), we also calculated model skill following Willmott (1981):

$$Skill = 1 - \frac{\sum |X_{mod} - X_{obs}|^2}{\sum (|X_{mod} - \overline{X_{obs}}| + |X_{obs} - \overline{X_{obs}}|)^2}, \quad (3)$$

where X_{obs} and X_{mod} are the observed and modeled variables, respectively, with the overbar indicating the time average. Skill provides an index of model-data agreement, with a skill of 1 indicating perfect agreement and 0 indicating complete disagreement. Skill has been widely used to evaluate the performance of numerical models (e.g., Warner et al., 2005). While the R^2 indicates the model's capability of capturing the seasonal trend and interannual variations, and RMSE indicates the overall misfit between model and observation, Skill can be regarded as a synthesis index to evaluate both the trend capturing and relative misfit.

3. Results

3.1. EOF Analysis

The first essential step in the proposed data-driven approach is to reduce the data-dimension. Through the EOF analysis and interpretation with our current knowledge of DO dynamics, we can understand the dominant processes controlling the DO variations. The spatial and temporal characteristics and the possible controlling mechanism of the first three EOF modes will be briefly described in the following paragraphs.

The first EOF mode is the dominant mode, accounting for 87% of the total variance. The map of the first mode is characterized with all positive values, meaning the changes in DO concentration are in phase among all the stations (Figures 5a–5b). The locations of higher values in the lower layer in middle to upper Bay correspond to the low DO region. The relative values in the map indicate the different magnitudes to which the DO concentration changes with time. For example, during the summer months when the temporal value of the first mode is negative, DO of the entire bay decreases from the long-term mean value and the maximum decrease occurs in the region with maximum map value (i.e., deep waters between 38°N and 39°N). The first mode has significant and obvious seasonality with the lowest temporal value in July, consistent with what is shown in Figure 2. This is a combined result of seasonal variation in water temperature, stratification, organic matter abundance, and water column respiration. Since these environmental factors share a high covariance with each other, it is not easy to identify which processes dominate the first mode.

It is necessary to point out that the first EOF mode not only reflects the seasonal characteristics of the DO dynamics but also contains strong signals of interannual variability. A simple linear comparison between July values of the first EOF mode and hypoxic volume calculated by Bever et al. (2013) shows that 62% of the interannual variations of July hypoxic volume during 1985–2008 can be explained by the first EOF mode (Figure 5c).

The map of the second EOF mode is characterized with opposite values between surface and subsurface layer in both middle and upper bay regions, suggesting this mode is controlled by mechanisms that have a different impact on DO between surface and subsurface layers. The opposite values between surface and subsurface is remarkable in the middle and upper bay regions (Figure 6b). The depth separating surface and subsurface is close to the long-term mean pycnocline depth, roughly at 10 m (Figure 6b). A seasonal cycle is noticeable in temporal variation with maximum mean value in May and maximum variability in April (Figure 6c); this pattern is similar to the chlorophyll-a concentration inside the bay (Figure 6d). Spring algae bloom in Chesapeake Bay occurs as early as the end of winter and the spring bloom usually reaches its maximum intensity around April (Harding, 1994). As a result, the variability of chlorophyll-a concentration also has the largest variability around April (Du & Shen, 2015). Algal bloom has a distinctly different impact on bottom and surface DO, with increasing surface DO due to photosynthesis and decreasing bottom DO due to the decomposition of settled organic matter. Linear regression analysis between the second EOF mode and chlorophyll-a concentration in the previous month shows a strong correlation, with R^2 of 0.35. It suggests a 1-month time-lag response of the second EOF mode to the chlorophyll-a concentration. We will show later that using the data-driven model, the R^2 can be improved significantly to 0.95 for the second EOF mode when multiple forcings are used.

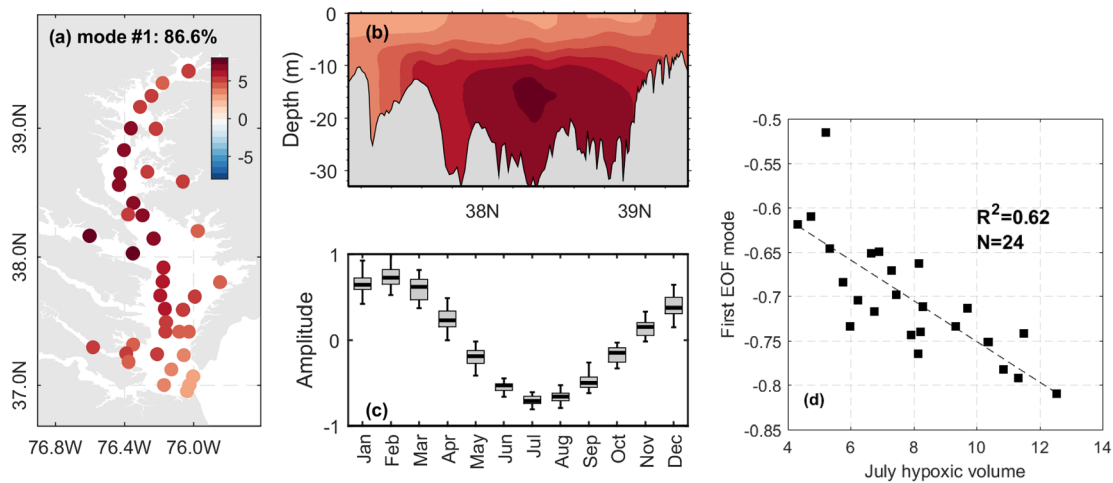


Figure 5. Characteristics of the first EOF mode that constitutes 86.6% of the total variance. (a) Horizontal distribution of bottom value; (b) the vertical distribution of the mode along the bay's main axis. Color dots in (a) and filled contours in (b) share the same color scale. (c) Box plots showing the seasonality of the time series. (d) Relationship between interannual variations of July hypoxic volume (data from Bever et al., 2013) and the July value of temporal variation of the first EOF mode.

The third EOF mode is also characterized with opposite values between surface and subsurface but in the region of lower to middle bay, with maximum opposition around 37.8°N where there is dramatic topographic change (Figure 7b). What is unique in the third EOF mode is that it has two peaks in its temporal value (Figure 7c). It is not clear what mechanisms are responsible to cause such unique spatial and temporal pattern. Not all EOF modes can be easily explained with our current knowledge. Sometimes, multiple mechanisms instead of a single one are responsible. In such cases, nonlinear models or machine learning techniques will be more suitable to explain the temporal variations. This is also why we try to use advanced deep-learning methods.

3.2. Simulation of DO

For the DO simulation, we trained the first five primary EOF modes. Little differences are found for the predicted DO in the testing period when the number of modes changes from five to nine. The first five modes

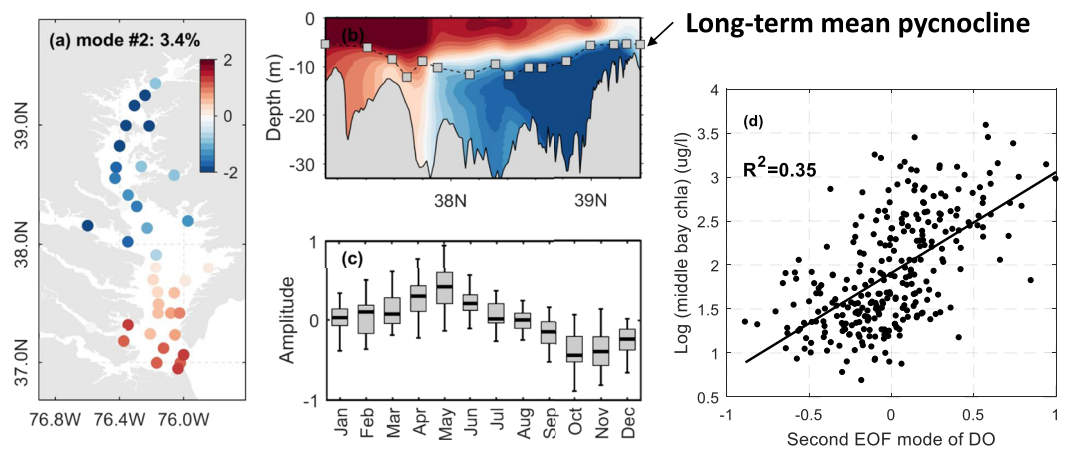


Figure 6. Characteristics of the second EOF mode. (a) Horizontal map of the bottom value. (b) Vertical distribution of second EOF mode along the mainstem, with gray solid rectangles denoting the long-term mean depth of pycnocline (determined as the depth where maximum salinity gradient occurs). (c) Seasonality of second EOF mode. (d) Relationship between chlorophyll-a averaged over middle bay stations, with each dot denoting the temporal value of second EOF mode at a given month and the logarithm of chlorophyll-a concentration in the previous month (i.e., with a 1-month time-lag).

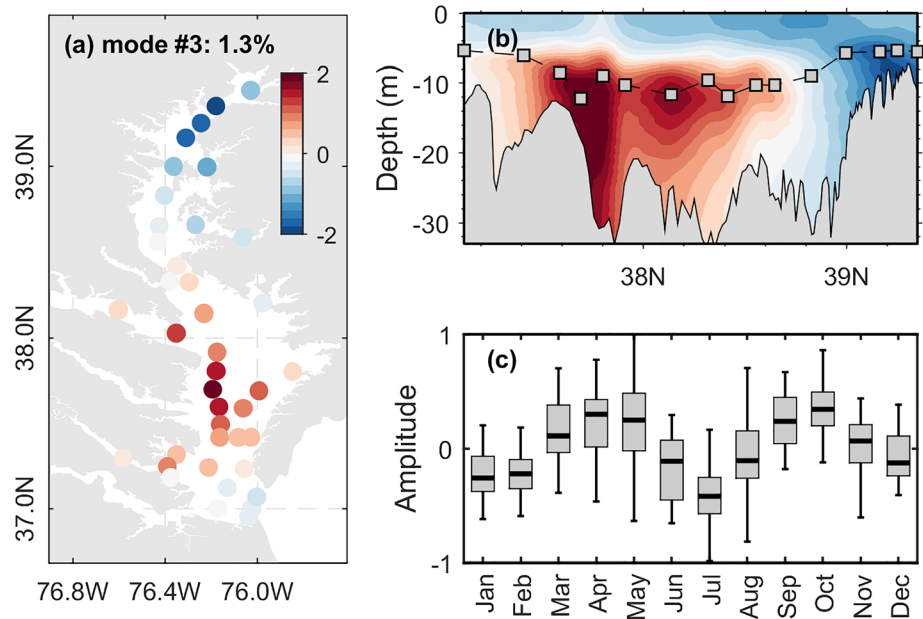


Figure 7. Characteristics of the third EOF mode (see caption of Figures 5a–5c).

account for 93% of the variance, while the first nine modes account for 95%. Including modes with minor contribution is believed unnecessary and may even introduce more noise to confound the major signals. From the training perspective, all of the five modes are well trained, with R^2 larger than 0.90 and skill over 0.95 (Figures 8a–8e). It is worth noting that the uncertainties (indicated by the standard deviation of the 100 neural network models' predictions) are almost negligible for the first mode and increase for later modes.

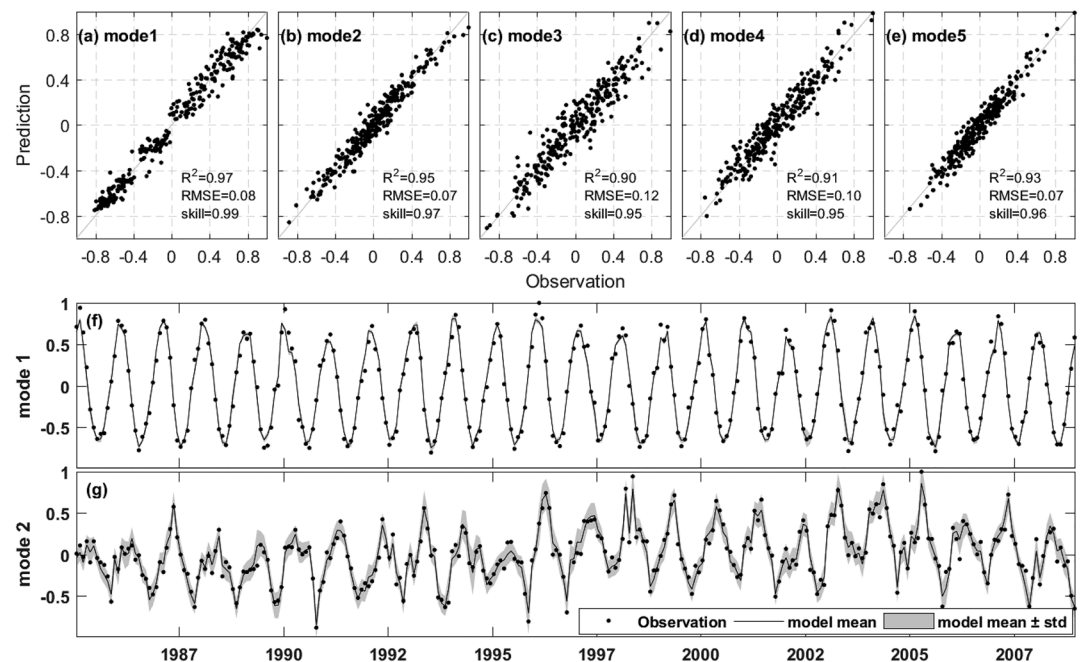


Figure 8. (a–e) Model training results for each of the first five EOF modes, with R^2 , RMSE, and model skill shown in the bottom right of each panel. (f–g) Observed and modeled time series of first two EOF modes in the training data set, with gray shade indicating the standard deviation of 100 times of neural network training.

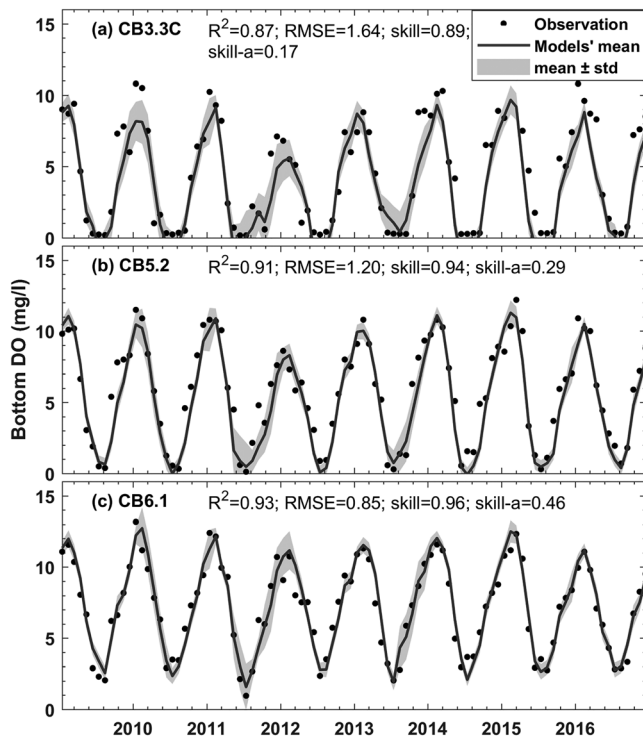


Figure 9. Observation (dot) and model prediction (line) of bottom DO at three selected mainstem stations, (a) CB3.3C, (b) CB5.2, and (c) CB6.1. Statistical indexes for the model performance are shown in the text on top. Skill-a indicates the skill for the anomaly prediction. The depth of the three stations are 25, 30, and 12 m, respectively. Gray shade indicates the standard deviation of 100 times of neural network training.

The performance of the model is evaluated from multiple aspects. First, we compared the bottom DO at each station between modeled and observed values. Compared to surface DO that is merely controlled by the air-sea exchange and water temperature, the bottom DO is affected by many more factors (e.g., stratification, organic matter decay, and sediment oxygen demand) and usually more difficult to simulate. Taking three mainstem stations (CB3.3C, CB5.2, and CB6.1, representing the upper, middle, and lower bay) as an example, the RMSE ranges from 0.85 to 1.64 mg/L (Figure 9). The peaks of bottom DO varied from year to year and their variabilities are captured by the model. For instance, the peak of bottom DO at CB3.3C in early 2012 was relatively small (less than 10 mg/L), compared to other years (Figure 9a). In particular, there was a sharp decrease of bottom DO in the fall of 2011, which was believed to be caused by the large freshwater input due to Hurricane Irene and the subsequent Tropical Storm Lee (Ye et al., 2019). The sharp decrease was noticeable at station CB3.3C, which is clearly captured by the model. Additionally, we calculated the model skill for the anomaly (i.e., the difference from the seasonal cycle) to examine the model performance in reproducing the deviations from seasonal cycle. The anomaly skill can be up to 0.46 (e.g., at station CB6.1); clearly, more efforts are needed to improve the anomaly prediction such as including more modes. The relatively low skill in anomaly prediction is a trade-off when including the entire signal of the target variable during the training stage. If one would like to focus on the interannual variability (say July only), one can train a model for that given month only, which in turn may raise another issue concerning the limited length of the training data.

The model performance at all of the 40 stations is statistically illustrated by the Taylor diagram (Figure 10). The correlation coefficients at the majority of the 40 stations from the proposed data-driven model concentrate in 0.90–0.95. Only one station has a correlation coefficient less than 0.8; this station (ET4.2) is located in Chester River, a small tributary discharging into the upper bay. It is likely that the bottom DO poorly

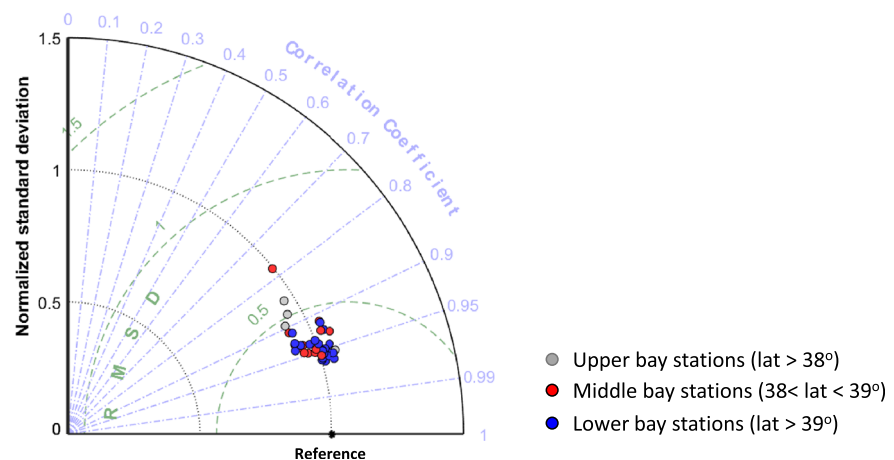


Figure 10. Taylor diagram illustrating the model performance for each of 40 stations, with different colors indicating stations in different regions. Only the 8-year testing data set (2009–2016) is used for the analysis. The radial distance from the origin is proportional to the ratio standard deviations; the azimuthal angle indicates the Pearson correlation coefficient; and the distance between each filled marker and the “reference” point indicates the centered root-mean-square deviation (RMSD).

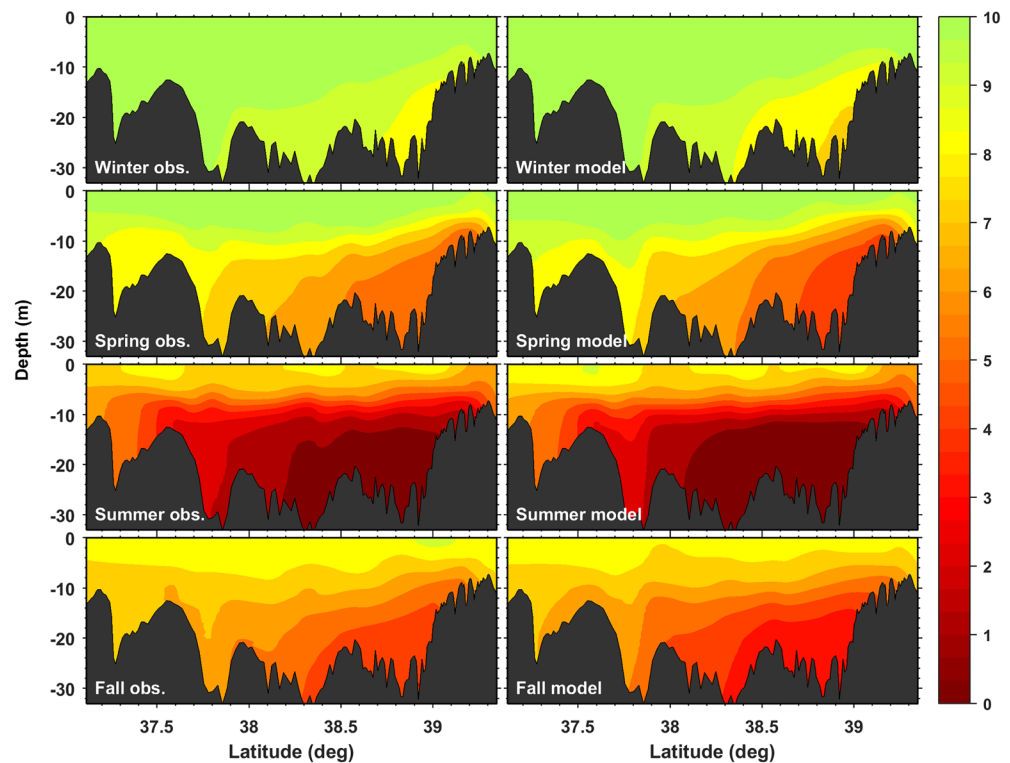


Figure 11. Spatial comparison between observed (left panels) and modeled (right panels) DO along the bay's mainstem.

simulated at this station is more influenced by local processes. Overall, the model performance is comparable to previous deterministic 3-D water quality models (e.g., Irby et al., 2016; Testa et al., 2014). For example, using a physical-biogeochemical model, Testa et al. (2014) shows the correlation coefficient of 0.80–0.95 for the bottom DO at mainstem stations (see the Taylor diagram Figure 6 in Testa et al., 2014). Irby et al. (2016) compared nine numerical models in simulating the bottom DO concentration, and, using a similar Taylor diagram (see Figure 8 in Irby et al., 2016), they show that the correlation coefficient from different models ranges from 0.8 to 0.9. With respect to the variance, the data-driven model predictions are close to the observation, with a mean, minimum, and maximum normalized standard deviation of 0.99, 0.92, and 1.07 over the 40 stations. The root-mean-square deviation (the third axes in the Taylor diagram) is around 0.3, which is smaller than some numerical model simulations (e.g., Testa et al., 2014). It is worth noting that only model performance of the bottom DO is presented here as it is typically more difficult to simulate accurately compared to the surface DO.

In terms of the spatial variations, the model performance is acceptable. As expected, the marked seasonal variations of the spatial distribution are well predicted for the testing period (Figure 11), despite the fact that the model seems to slightly underestimate the bottom DO. Furthermore, by comparing the spatial distribution of summer DO in the testing period, it is obvious that the model captures the overall change of hypoxic area (Figure 12). From year to year, the summer hypoxic area changed, with a minimum area in 2012 (2 mg/L contour starts from 38°N) and a maximum area in 2011 (2 mg/L contour starts from 37.6°N). Observation and model predictions are consistent in the distribution of 2 and 4 mg/L contours along the mainstem. It is expected that there is still noticeable bias especially concerning the severely hypoxic area (e.g., area with DO less than 1 mg/L), which is also a problem for previous models, either empirical or numerical ones (e.g., Cerco & Noel, 2013; Irby et al., 2016; Testa et al., 2014). One of the reasons is that there is no signal in DO when DO becomes zero, while forcing variables are still varying, resulting in high uncertainty at these anoxic regions.

It is worth noting that in the proposed approach, the EOF spatial patterns extracted from the training data set are assumed unchanged during the testing period. This assumption is valid if the underlying

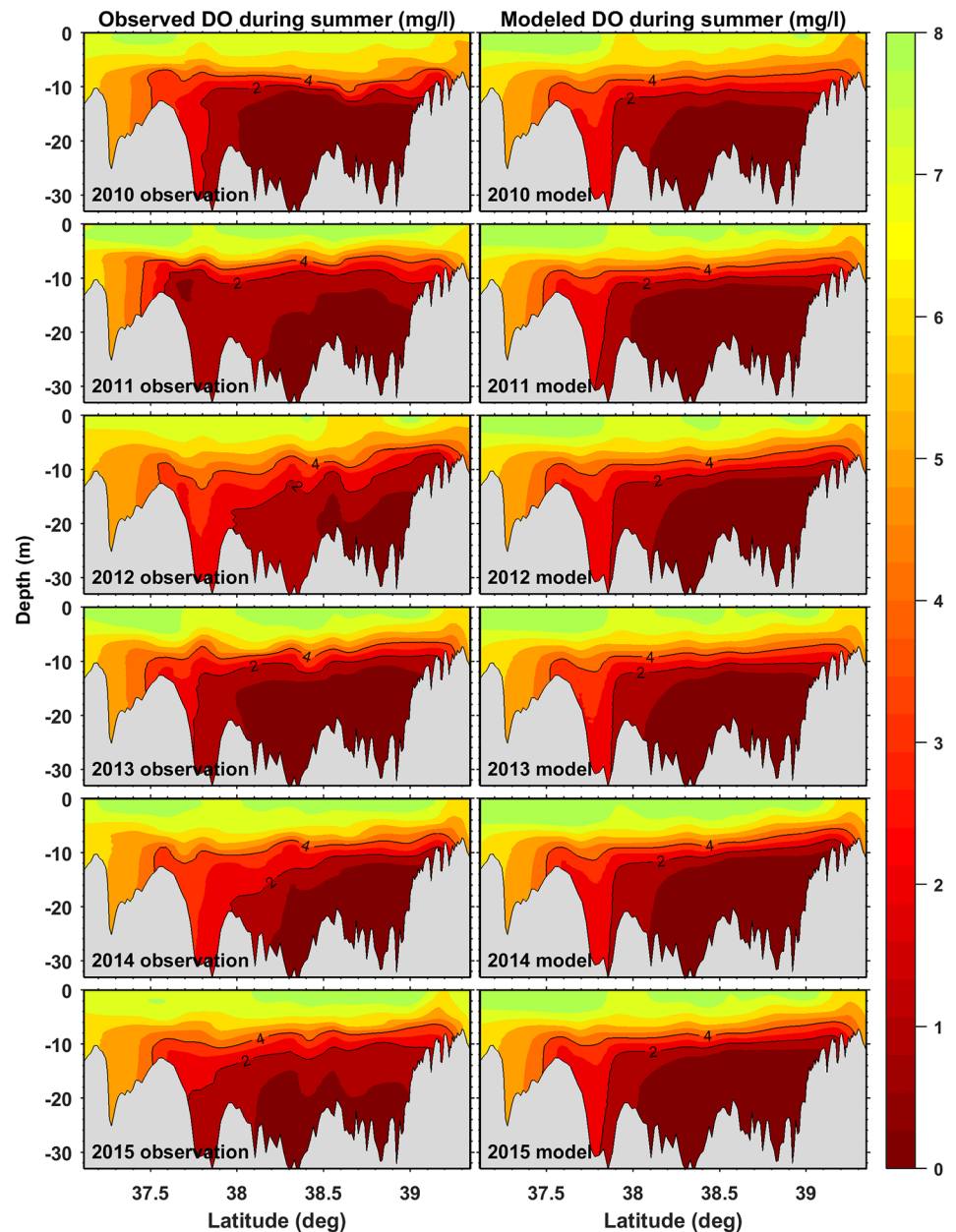


Figure 12. Vertical profile of the summer-mean DO concentration (averaged over June–August) along the bay's mainstem, with black lines denoting contour of 2 and 4 mg/L. Only testing data set are shown here in order to demonstrate the model's capability in reproducing the spatial distribution when external forcings are provided.

physical and biological processes have not undergone dramatic change, which is true except for some rare cases (e.g., when there is a regime shift) in which the covariance among different stations changes dramatically. In the case of DO in Chesapeake Bay, additional EOF analysis using the DO data in the testing period confirms the validity of the assumption. The spatial patterns for the first three EOF modes in the testing period (Figures S3–S5) are nearly identical to those in the training period (Figures 5–7) despite the little differences, which can be attributed to the shorter records in the testing period. For practical application, both EOF analysis and training can be conducted dynamically as more data will become available in future to ensure the test period has the same spatial pattern as the training period.

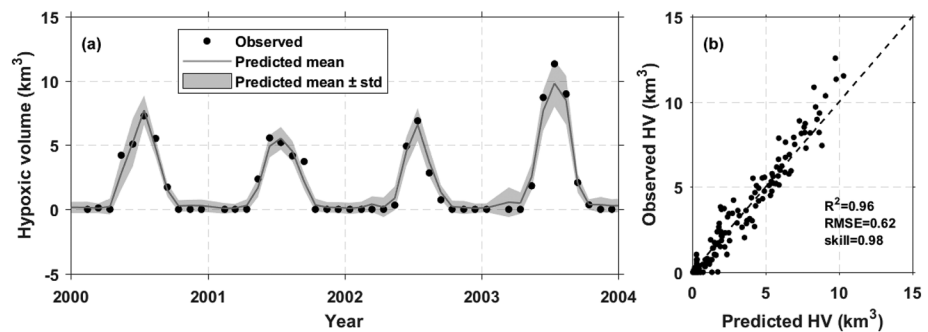


Figure 13. Training part for the hypoxic volume simulation. (a) shows the time series for the last four year in training period, while (b) shows the monthly data for the entire training period. Only training data set of 1985–2004 is used for the training. In (a), the gray shading indicates the uncertainties calculated as the standard deviation of results from 100 times of neural network training.

3.3. Simulation of Hypoxic Volume

Similar methodology was applied to simulate the hypoxic volume in the Chesapeake Bay. Here we used the 28-year hypoxic volume time series (1985–2012) calculated by Bever et al. (2013) as the target variable. The hypoxic volume is a scalar index used to quantify overall DO condition in the Bay. Different from DO simulation, only one time series is used as the target variable and there is no need to use EOF analysis. The forcing selection and transformation and the model training algorithm applied are the same as for DO simulation. Input forcings include freshwater discharges and nutrient loads from major rivers, air temperature, and wind. For the training part, the model yields R^2 of 0.96 and $RMSE = 0.62 \text{ km}^3$ (Figure 13).

We also conducted tests using other machine learning methods including multiple linear regression and decision tree with the same input forcing as used in the trained neural network models. Results show that the linear model ($RMSE = 1.46 \text{ km}^3$) is unable to capture the interannual variability (especially during summer months) and it frequently produces negative value during the winter months (Figure 14). On the contrary, the decision tree ($RMSE = 1.35 \text{ km}^3$) and neural network ($RMSE = 1.29 \text{ km}^3$) methods seem to have better performance. Particularly, the decision tree almost perfectly predicts the zero values. While it is hard to determine which method is always the best, nonlinear models are usually regarded as more suitable than linear models when dealing with water quality variables. It is often a good strategy to test multiple data methods and examine the performance difference among them. Using the ensemble mean of results from multiple methods could be a good option to account for the uncertainties associated with choice of the machine learning method.

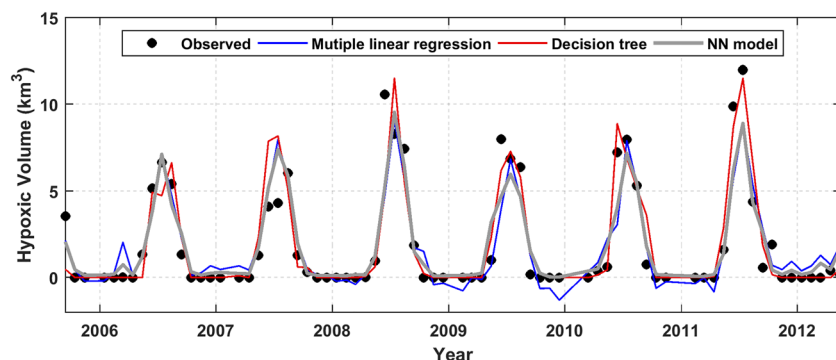


Figure 14. Testing part for the hypoxic volume simulation. Besides the neural network model, other methods including multiple linear regression, decision tree, and Bayesian regression are tested using the same input as in neural network.

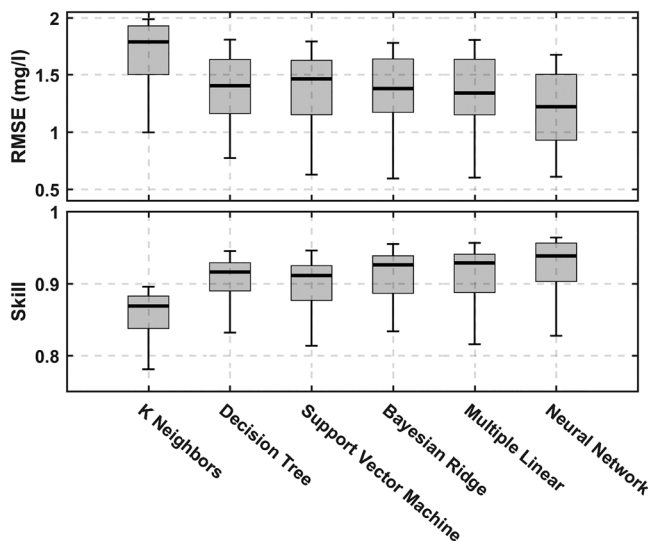


Figure 15. The model performance indicated by root-mean-square error (RMSE) and skill for the bottom DO. The gray bars and error bars indicate the mean and standard deviation of the performance over the 40 stations.

the selected primary EOF modes. The high computational efficiency makes the data-driven model an efficient way for environmental research and management. Since the model uses external forcings as used by a traditional 3-D water quality model, the data-driven approach can be used to predict water quality conditions under future climate scenarios in response to change in environmental forcing conditions, such as changes in temperature, wind, and river flow. The nonlinear influence of forcing conditions on water quality is implicitly inherited inside the data-driven model (Shen et al., 2019). However, it should be noted that such predictions may only be possible when the training data set has included the variations of specific forcing and that the trained model has correctly resolved the response of target variable to the given forcing. Finally, even though the approach is used for the DO problem in Chesapeake Bay, similar or same methodology is readily applicable to other coastal systems, as long as there is enough data to train a model with reasonable accuracy. How much data are enough? There is not a straight answer. The required data length depends on the problem to be studied and the timescale of major driving processes for both the biogeochemical and physical condition. The training data set should at least cover multiple dominant periods during which major signals (e.g., seasonal cycle and interannual variations) are noticeable.

It is important to point out that there are a variety of data methods well developed for machine learning, ranging from the simplest linear regression to sophisticated nonlinear regression such as K nearest neighbors, random tree, and support vector machine (Figure S6). We tested these methods for the DO problem with the same input variables and target variable. Except for the K nearest neighbors, all other methods tested yield similar model performance, even though the neural network model appears to be slightly better than others (Figure 15).

4.2. Lessons Learned

When using the model, one has to be cautious when collecting and transforming forcing data. As described in section 2, in this study, only forcing data continuously monitored at boundaries of the systems are used. These boundaries can be the air-sea interface, river boundary, and estuary-ocean interface. No data within the water body is used. For instance, the nutrient concentration and water salinity at mainstem stations are not used, even though they are available and believed to be highly correlated with the DO condition. Such protocol makes the proposed method distinctly different from many previous studies that used the in situ measurement to represent either the physical or biogeochemical conditions (e.g., Ross & Stock, 2019; Scavia et al., 2006). For instance, to estimate the hypoxic volume, Scavia et al. (2006) used the bottom-surface difference of DO to calculate the vertical DO flux. Scardi and Harding (1999) used in situ measured parameters including chlorophyll concentration, salinity, and photic zone depth to train a

4. Discussion

4.1. Robustness of Data-Driven Model

Different from traditional empirical or numerical models, the data-driven model proposed in this study has shown advantages in several aspects. First, the approach is purely based on reliable and systematic monitoring or reanalysis data, which avoids the uncertainties propagating and accumulating from base level of hydrodynamic models to higher level of water quality models. The error accumulation is a common issue for many deterministic model systems. With advanced numerical tools and data-assimilation techniques, performance of hydrodynamic simulation has been improved greatly over time. Nevertheless, errors still exist even in a high-resolution numerical hydrodynamic models with a high order of numerical schemes (e.g., Irby et al., 2016; Testa et al., 2014; Ye et al., 2019). Second, the proposed approach can simulate not only the temporal but also spatial variations by extracting the major components of the target water quality variable. This is very different from previous 1-D empirical models (e.g., Scardi & Harding, 1999; Scavia et al., 2006). Third, the approach is highly computationally efficient, taking about 1 hr with one cpu to train the models for all

neural-network model of phytoplankton primary production. Avoiding using such type of forcings is crucial to make the data-driven model more capable for environmental management.

Another important lesson learned for the data-driven model is to choose appropriate transformations for each forcing. For coastal systems, especially large coastal systems, delayed response of water quality to change of external forcing is very common. For instance, Malone et al. (1988) suggest the maximum summer productivity in the mesohaline reach of Chesapeake Bay is caused by the recycled nutrient delivered to the system during the previous spring. From numerical experiments, Lee et al. (2013) suggests winter-spring wind could affect the summer hypoxia (i.e., with a time-lag of 4 months) by affecting the circulation pattern and thus the transport of spring bloom biomass along the mainstem and between shoal and deep channel. The time-lag shall likely differ in different regions of a coastal system and for different processes. It is usually subjective in determining how long the time-lag is. The ASFT tool proposed in this study will ease this effort and is particularly useful when training multiple models using a variety of forcings.

We found that wind is one of the important variables for hypoxic volume simulation and the wind speed is first selected forcing by the ASFT module (Table S6). Although wind forcing is considered to only have a short-term impact, wind forcing can be more influential than expected for interannual dynamics. Recent studies confirm the great control of wind on stratification, estuarine circulation, and thus hypoxic volume (Du & Shen, 2016; Du et al., 2018; Jiang & Xia, 2017; Scully, 2010; Shen et al., 2013; Wilson et al., 2015). Numerical experiments in previous studies (e.g., Hong & Shen, 2013; Scully, 2013) suggest the vertical mixing is less sensitive to the synoptic variability of river discharge but more sensitive to the wind condition. Scully (2013, 2016) finds that the mean summer wind speed is the single-most important physical variable contributing to the variations of hypoxic volume. Changes in physical condition induced by changing wind fields in the future should be considered for the environmental and water quality management in Chesapeake Bay.

The selection of forcing and transformation shall not be used directly to interpret the relative importance of one forcing over another because of high covariance among different forcings and the difficulty to isolate the influence of each forcing based on the correlation analysis. If two forcings are highly correlated, only the forcing with higher R^2 is selected, which is different from deterministic models. Nevertheless, the information of forcing selection is helpful to determine the relative complexity of each mode. For instance, only four forcings are selected for the first EOF mode (Table S1) while more than 20 forcings for second–fifth modes, suggesting higher complexity of dynamics responsible for these later modes.

Of particular interest is the forcing selection for the first EOF mode, the dominant mode containing not only the seasonal signal of bottom DO but also the interannual signal of bay-wide hypoxia (Figure 5). The first mode is only related to air temperature, Susquehanna streamflow shifted by 40 days, and downwelling short-wave radiation shifted by 10 days (Table S1). Considering the consistency of seasonality among the first mode, air temperature, and solar radiation, it seems that the data model uses radiation and/or temperature to learn the seasonal cycle, and uses the streamflow in controlling interannual variability. Because nutrient loadings are highly correlated to streamflow, the information of nutrient loading is implicitly included. This is broadly consistent with other studies using either statistic or 3-D numerical approaches (e.g., Hagy et al., 2004; Scavia et al., 2006; Testa et al., 2014).

4.3. Limitations of the Data-Driven Model

It is important to acknowledge not only the advantages but also the limitations of a method. First, the accuracy of a trained model depends greatly on the length of data records. Even in a heavily monitored coastal system such as Chesapeake Bay, the length of data is still limited because of the monitoring frequency. One critical question is whether results of the bi-monthly surveys is representative for the monthly mean condition. By comparing the variability of measurement and simulation results of DO in Chesapeake Bay, Bever et al. (2013) demonstrated the monthly measurements need to be corrected in order to be spatially and temporally representative. Without enough temporal resolution, the EOF analysis of the existing data set may be overwhelmed by minor modes, requiring more modes to be included in order to cover the major variance. Second, the model performance may be hampered by the unavailability of systematic forcing. Monitoring instrument malfunction occurs frequently, especially after major extreme weather events such as hurricanes, which may result in a significant data gap. In these cases, reanalysis model results that

assimilate observations could be good data sources. Taking the wind data for instance, instead of using the monitoring data, we used the outputs of reanalysis atmospheric models. Perhaps, the most concerning shortcoming is that data-driven models are not process-orientated, making them hard for process-based management and research. As many forcing variables are highly correlated with each other, the contribution of an individual forcing variable cannot be fully evaluated based on sensitivity tests although the model has a high predictive skill. However, it does not mean such models cannot be used for environmental management. Actually, there is an increasing need of using data-driven model for water quality management (e.g., Chang et al., 2015; Orouji et al., 2013; Ross & Stock, 2019; Shen et al., 2019). Shen et al. (2019) shows a data-driven model can be used for predicting the response of Chl-a to changes in nutrient loading if the appropriate parameter variables are used. Through model sensitivity tests by introducing different forcings and using different transformations, the data-driven model could be helpful in identifying possibly important factors.

5. Conclusions

We present a data-driven model to efficiently and accurately simulate and predict DO conditions in estuarine and coastal waters. Different from previous statistical models that often use in situ measurements, the proposed approach relies purely on the external forcings, which make it more suitable for environmental assessment. Evaluation of the model performance suggests a high capability of data-driven model for water quality simulation. Based on a similar forcing data set as required by a 3-D numerical model, the proposed approach predicts well the DO and hypoxic condition spatially and temporally. Even though the model is tested for the DO condition in Chesapeake Bay, the framework and methodology are readily transferable to any other coastal systems that are systematically monitored. Overall, this study provides a robust framework and methodology, upon which future research could be based.

With the quickly accumulated observational data and latest advances in machine-learning techniques, the data-driven model is a promising approach with high efficiency for water quality modeling and environmental management in the near future. In fact, there is increasing interest in using machine-learning techniques for water quality simulations (e.g., Ross & Stock, 2019; Shen et al., 2019). There are, however, still some questions that remain to be further explored. For example, how to transform the forcing and target variable appropriately? How to faithfully extract the principal component? And how to determine whether noncontinuous measurements are representative for monthly or daily conditions? Even though this study has attempted to answer them, the answers may vary depending on specific problems and characteristics of the given estuarine and coastal system.

Data Availability Statement

The long-term monitoring data for the dissolved oxygen in Chesapeake Bay are available at Chesapeake Bay Program (<http://data.chesapeakebay.net/WaterQuality>). Freshwater discharge data of major rivers are available United States Geological Survey (Susquehanna River at https://waterdata.usgs.gov/md/nwis/uv?site_no=01578310; Potomac River at https://waterdata.usgs.gov/md/nwis/uv?site_no=01646500; James River at <https://waterdata.usgs.gov/usa/nwis/uv?02037500>; Choptank River at <https://waterdata.usgs.gov/usa/nwis/uv?01491000>). Data of air temperature at Chesapeake Bay Bridge Tunnel are available at NOAA (<https://tidesandcurrents.noaa.gov/stationhome.html?id=8638863>). Hourly reanalysis wind data are produced by European Centre for Medium-Range Weather Forecasts and available online (<https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5>). The code showing the framework of the data method can be found online (<https://zenodo.org/record/3973774>).

References

- Bajo, M., & Umgiesser, G. (2010). Storm surge forecast through a combination of dynamic and neural network models. *Ocean Modelling*, 33(1–2), 1–9. <https://doi.org/10.1016/j.ocemod.2009.12.007>
- Bever, A. J., Friedrichs, M. A. M., Friedrichs, C. T., Scully, M. E., & Lanerolle, L. W. J. (2013). Combining observations and numerical model results to improve estimates of hypoxic volume within the Chesapeake Bay, USA. *Journal of Geophysical Research: Oceans*, 118, 4924–4944. <https://doi.org/10.1002/jgrc.20331>
- Bianchi, T. S., DiMarco, S. F., Cowan, J. H., Hetland, R. D., Chapman, P., Day, J. W., & Allison, M. A. (2010). The science of hypoxia in the northern Gulf of Mexico: A review. *Science of the Total Environment*, 408(7), 1471–1484. <https://doi.org/10.1016/j.scitotenv.2009.11.047>

Acknowledgments

We want to thank the three anonymous reviewers for their valuable and helpful comments. This is contribution No. 3934 of the Virginia Institute of Marine Science, College of William and Mary.

- Breitburg, D., Levin, L. A., Oschlies, A., Grégoire, M., Chavez, F. P., Conley, D. J., et al. (2018). Declining oxygen in the global ocean and coastal waters. *Science*, 359, eaam7240. <https://doi.org/10.1126/science.aam7240>
- Campolo, M., Andreussi, P., & Soldati, A. (1999). River flood forecasting with a neural network model. *Water Resources Research*, 35(4), 1191–1197. <https://doi.org/10.1029/1998WR900086>
- Carstensen, J., Andersen, J. H., Gustafsson, B. G., & Conley, D. J. (2014). Deoxygenation of the Baltic Sea during the last century. *Proceedings of the National Academy of Sciences*, 111(15), 5628–5633. <https://doi.org/10.1073/pnas.1323156111>
- Cerco, C. F., & Noel, M. R. (2013). Twenty-one-year simulation of Chesapeake Bay water quality using the CE-QUAL-ICM eutrophication model. *Journal of the American Water Resources Association*, 49, 1119–1133. <https://doi.org/10.1111/jawr.12107>
- Chang, F., Tsai, Y., Chen, P., Coynel, A., & Vachaud, G. (2015). Modeling water quality in an urban river using hydrological factors—Data driven approaches. *Journal of Environmental Management*, 151, 87–96. <https://doi.org/10.1016/j.jenvman.2014.12.014>
- Chang, F. J., & Chen, Y. C. (2003). Estuary water-stage forecasting by using radial basis function neural network. *Journal of Hydrology*, 270(1–2), 158–166. [https://doi.org/10.1016/S0022-1694\(02\)00289-5](https://doi.org/10.1016/S0022-1694(02)00289-5)
- Conley, D. J., Humborg, C., Rahm, L., Savchuk, O. P., & Wulff, F. (2002). Hypoxia in the Baltic Sea and basin-scale changes in phosphorus biogeochemistry. *Environmental Science & Technology*, 36(24), 5315–5320. <https://doi.org/10.1021/es025763w>
- Da, F., Friedrichs, M. A. M., & St-Laurent, P. (2018). Impacts of atmospheric nitrogen deposition and coastal nitrogen fluxes on oxygen concentrations in Chesapeake Bay. *Journal of Geophysical Research: Oceans*, 123, 5004–5025. <https://doi.org/10.1029/2018JC014009>
- Deng, J., Paerl, H. W., Qin, B., Zhang, Y., Zhu, G., Jeppesen, E., et al. (2018). Climatically-modulated decline in wind speed may strongly affect eutrophication in shallow lakes. *Science of the Total Environment*, 645, 1361–1370. <https://doi.org/10.1016/j.scitotenv.2018.07.208>
- Diaz, R. J., & Rosenberg, R. (2008). Spreading dead zones and consequences for marine ecosystems. *Science*, 321(5891), 926–929. <https://doi.org/10.1126/science.1156401>
- Du, J., Park, K., Shen, J., Zhang, Y. J., Yu, X., Ye, F., et al. (2019). A hydrodynamic model for Galveston Bay and the shelf in the northern Gulf of Mexico. *Ocean Science*, 15(4), 951–966. <https://doi.org/10.5194/os-15-951-2019>
- Du, J., & Shen, J. (2015). Decoupling the influence of biological and physical processes on the dissolved oxygen in the Chesapeake Bay. *Journal of Geophysical Research: Oceans*, 118, 3114–3127. <https://doi.org/10.1002/jgrc.20224>
- Du, J., & Shen, J. (2016). Water residence time in Chesapeake Bay from 1980–2012. *Journal of Marine Systems*, 164, 101–111. <https://doi.org/10.1016/j.jmarsys.2016.08.011>
- Du, J., Shen, J., Park, K., Wang, Y. P., & Yu, X. (2018). Worsened physical condition due to climate change contributes to the increasing hypoxia in Chesapeake Bay. *Science of the Total Environment*, 630, 707–717. <https://doi.org/10.1016/j.scitotenv.2018.02.265>
- Fennel, K., & Testa, J. M. (2019). Biogeochemical controls on coastal hypoxia. *Annual Review of Marine Science*, 11(1), 105–130. <https://doi.org/10.1146/annurev-marine-010318-095138>
- Hagan, M. T., & Menhaj, M. (1999). Training feed-forward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*, 5, 989–993.
- Hagy, J. D., Boynton, W. R., Keefe, C. W., & Wood, K. V. (2004). Hypoxia in Chesapeake Bay, 1950–2001: Long-term change in relation to nutrient loading and river flow. *Estuaries*, 27(4), 634–658. <https://doi.org/10.1007/BF02907650>
- Harding, L. W. (1994). Long-term trends in the distribution of phytoplankton in Chesapeake Bay: Roles of light, nutrients and streamflow. *Marine Ecology Progress Series*, 104, 267–291. <https://doi.org/10.3354/meps104267>
- Hong, B., & Shen, J. (2012). Responses of estuarine salinity and transport processes to potential future sea-level rise in the Chesapeake Bay. *Estuarine, Coastal and Shelf Science*, 104–105, 33–45. <https://doi.org/10.1016/j.ecss.2012.03.014>
- Hong, B., & Shen, J. (2013). Linking dynamics of transport timescale and variations of hypoxia in the Chesapeake Bay. *Journal of Geophysical Research: Oceans*, 118, 6017–6029. <https://doi.org/10.1002/2013JC008859>
- Hsu, K., Gupta, H. V., & Sorooshian, S. (1995). Artificial neural network modeling of the rainfall-runoff process. *Water Resources Research*, 31(10), 2517–2530. <https://doi.org/10.1029/95WR01955>
- Irby, I. D., Friedrichs, M. A. M., Friedrichs, C. T., Bever, A. J., Hood, R. R., Lanerolle, W. J., et al. (2016). Challenges associated with modeling low-oxygen waters in Chesapeake Bay: A multiple model comparison. *Biogeosciences*, 13(7), 2011–2028. <https://doi.org/10.5194/bg-13-2011-2016>
- Jiang, L., & Xia, M. (2017). Wind effects on the spring phytoplankton dynamics in the middle reach of the Chesapeake Bay. *Ecological Modeling*, 363, 68–80. <https://doi.org/10.1016/j.ecolmodel.2017.08.026>
- Karstensen, J., Stramma, L., & Visbeck, M. (2008). Oxygen minimum zones in the eastern tropical Atlantic and Pacific oceans. *Progress in Oceanography*, 77(4), 331–350. <https://doi.org/10.1016/j.pocan.2007.05.009>
- Keiner, L. E., & Yan, X. H. (1998). A neural network model for estimating sea surface chlorophyll and sediments from thematic mapper imagery. *Remote Sensing of Environment*, 66(2), 153–165. [https://doi.org/10.1016/S0034-4257\(98\)00054-6](https://doi.org/10.1016/S0034-4257(98)00054-6)
- Kemp, W. M., Boynton, W. R., Adolf, J. E., Boesch, D. F., Boicourt, W. C., & Brush, G. (2005). Eutrophication of Chesapeake Bay: Historical trends and ecological interactions. *Marine Ecology Progress Series*, 303, 1–29. <https://doi.org/10.3354/meps303001>
- Krasnopolsky, V. M. (2007). Neural network emulations for complex multidimensional geophysical mappings: Applications of neural network techniques to atmospheric and oceanic satellite retrievals and numerical modeling. *Reviews of Geophysics*, 45, RG3009. <https://doi.org/10.1029/2006RG000200>
- Lee, Y. J., Boynton, W. R., Li, M., & Li, Y. (2013). Role of late winter–spring wind influencing summer hypoxia in Chesapeake Bay. *Estuaries and Coasts*, 36(4), 683–696. <https://doi.org/10.1007/s12237-013-9592-5>
- Malone, T., Kemp, W., Ducklow, H., Boynton, W., Tuttle, J., & Jonas, R. (1986). Lateral variation in the production and fate of phytoplankton in a partially stratified estuary. *Marine Ecology Progress Series*, 32, 149–160. <https://doi.org/10.3354/meps032149>
- Malone, T. C., Crocker, L. H., Pike, S. E., & Wendler, B. W. (1988). Phytoplankton production in a partially stratified estuary. *Marine Ecology Progress Series*, 48, 235–249. <https://doi.org/10.3354/meps048235>
- Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, 11(2), 431–441. <https://doi.org/10.1137/0111030>
- Meier, H. E. M., Hordoir, R., Andersson, H. C., Dieterich, C., Eilola, K., Gustafsson, B. G., et al. (2012). Modeling the combined impact of changing climate and changing nutrient loads on the Baltic Sea environment in an ensemble of transient simulations for 1961–2099. *Climate Dynamics*, 39(9–10), 2421–2441. <https://doi.org/10.1007/s00382-012-1339-7>
- Morris, J. T., Porter, D., Neet, M., Noble, P. A., Schmidt, L., Lapine, L. A., & Jensen, J. R. (2005). Integrating LIDAR elevation data, multi-spectral imagery and neural network modelling for marsh characterization. *International Journal of Remote Sensing*, 26(23), 5221–5234. <https://doi.org/10.1080/01431160500219018>
- Muller, A. C., & Muller, D. L. (2015). Forecasting future estuarine hypoxia using a wavelet based neural network model. *Ocean Modelling*, 96, 314–323. <https://doi.org/10.1016/j.ocemod.2015.11.003>

- Murphy, R. R., Kemp, W. M., & Ball, W. P. (2011). Long-term trends in Chesapeake Bay seasonal hypoxia, stratification, and nutrient loading. *Estuaries and Coasts*, 34(6), 1293–1309. <https://doi.org/10.1007/s12237-011-9413-7>
- Muttill, N., & Chau, K. W. (2006). Neural network and genetic programming for modelling coastal algal blooms. *International Journal of Environment and Pollution*, 28(3/4), 223. <https://doi.org/10.1504/IJEP.2006.011208>
- Newcombe, C. L., & Horne, W. A. (1938). Oxygen-poor waters of the Chesapeake Bay. *Science*, 88(2273), 80–81. <https://doi.org/10.1126/science.88.2273.80>
- Orouji, H., Bozorg-haddad, O., & Fallah-Mehdipour, E. (2013). Modeling of water quality parameters using data-driven models. *Journal of Environmental Engineering*, 139, 947–957. [https://doi.org/10.1061/\(ASCE\)EE.1943-7870.0000706](https://doi.org/10.1061/(ASCE)EE.1943-7870.0000706)
- Paliwal, M., & Kumar, U. A. (2009). Neural networks and statistical techniques: A review of applications. *Expert Systems with Applications*, 36(1), 2–17. <https://doi.org/10.1016/j.eswa.2007.10.005>
- Rabalais, N. N., Turner, R. E., & Wiseman, W. J. (2002). Gulf of Mexico hypoxia, a.k.a. “the dead zone”. *Annual Review of Ecology and Systematics*, 33(1), 235–263. <https://doi.org/10.1146/annurev.ecolsys.33.010802.150513>
- Recknagel, F. (2001). Applications of machine learning to ecological modelling. *Ecological Modelling*, 146(1–3), 303–310. [https://doi.org/10.1016/S0304-3800\(01\)00316-7](https://doi.org/10.1016/S0304-3800(01)00316-7)
- Ross, A. C., & Stock, C. A. (2019). An assessment of the predictability of column minimum dissolved oxygen concentrations in Chesapeake Bay using a machine learning model. *Estuarine, Coastal and Shelf Science*, 221, 53–65. <https://doi.org/10.1016/j.ecss.2019.03.007>
- Scardi, M. (1996). Artificial neural networks as empirical models for estimating phytoplankton production. *Marine Ecology Progress Series*, 139(1–3), 289–299. <https://doi.org/10.3354/meps139289>
- Scardi, M., & Harding, L. W. (1999). Developing an empirical model of phytoplankton primary production: A neural network case study. *Ecological Modelling*, 120(2–3), 213–223. [https://doi.org/10.1016/S0304-3800\(99\)00103-9](https://doi.org/10.1016/S0304-3800(99)00103-9)
- Scavia, D., Kelly, E. L. A., & Hagy, J. D. (2006). A simple model for forecasting the effects of nitrogen loads on Chesapeake Bay hypoxia. *Estuaries and Coasts*, 29(4), 674–684. <https://doi.org/10.1007/BF02784292>
- Scully, M. E. (2010). Wind modulation of dissolved oxygen in Chesapeake Bay. *Estuaries and Coasts*, 33(5), 1164–1175. <https://doi.org/10.1007/s12237-010-9319-9>
- Scully, M. E. (2013). Physical controls on hypoxia in Chesapeake Bay: A numerical modeling study. *Journal of Geophysical Research: Oceans*, 118, 1239–1256. <https://doi.org/10.1002/jgrc.20138>
- Scully, M. E. (2016). The contribution of physical processes to inter-annual variations of hypoxia in Chesapeake Bay: A 30-yr modeling study. *Limnology and Oceanography*, 61(6), 2243–2260. <https://doi.org/10.1002/lno.10372>
- Shen, J., Hong, B., & Kuo, A. Y. (2013). Using timescales to interpret dissolved oxygen distributions in the bottom waters of Chesapeake Bay. *Limnology and Oceanography*, 58(6), 2237–2248. <https://doi.org/10.4319/lno.2013.58.6.2237>
- Shen, J., Qin, Q., Wang, J., & Sisson, M. (2019). A data-driven modeling approach for simulating algal blooms in the tidal freshwater of James River in response to riverine nutrient loading. *Ecological Modelling*, 398, 44–54. <https://doi.org/10.1016/j.ecolmodel.2019.02.005>
- Shen, J., & Wang, H. V. (2007). Determining the age of water and long-term transport timescale of the Chesapeake Bay. *Estuarine, Coastal and Shelf Science*, 74(4), 585–598. <https://doi.org/10.1016/j.ecss.2007.05.017>
- Taft, J. L., Taylor, W. R., Hartwig, E. O., & Loftus, R. (1980). Seasonal oxygen depletion in Chesapeake Bay. *Estuaries*, 3(4), 242–247. <https://doi.org/10.2307/1352079>
- Testa, J. M., Li, Y., Lee, Y. J., Li, M., Brady, D. C., Di Toro, D. M., et al. (2014). Quantifying the effects of nutrient loading on dissolved O₂ cycling and hypoxia in Chesapeake Bay using a coupled hydrodynamic – biogeochemical model. *Journal of Marine Systems*, 139, 139–158. <https://doi.org/10.1016/j.jmarsys.2014.05.018>
- Vaquier-Sunyer, R., & Duarte, C. M. (2008). Thresholds of hypoxia for marine biodiversity. *Proceedings of the National Academy of Sciences of the United States of America*, 105(40), 15,452–15,457. <https://doi.org/10.1073/pnas.0803833105>
- Vilas, L. G., Spyarakos, E., & Torres Palenzuela, J. M. (2011). Neural network estimation of chlorophyll a from MERIS full resolution data for the coastal waters of Galician rias (NW Spain). *Remote Sensing of Environment*, 115(2), 524–535. <https://doi.org/10.1016/j.rse.2010.09.021>
- Warner, J. C., Geyer, W. R., & Lerczak, J. A. (2005). Numerical modeling of an estuary: A comprehensive skill assessment. *Journal of Geophysical Research*, 110, C05001. <https://doi.org/10.1029/2004JC002691>
- Willmott, C. (1981). On the validation of models. *Physical Geography*, 2(2), 184–194. <https://doi.org/10.1080/02723646.1981.10642213>
- Wilson, R. E., Bratton, S. D., Wang, J., & Colle, B. A. (2015). Evidence for directional wind response in controlling inter-annual variations in duration and areal extent of summertime hypoxia in western Long Island sound. *Estuaries and Coasts*, 38(5), 1735–1743. <https://doi.org/10.1007/s12237-014-9914-2>
- Ye, F., Zhang, Y. J., He, R., Wang, Z., Wang, H. V., & Du, J. (2019). Third-order WENO transport scheme for simulating the baroclinic eddying ocean on an unstructured grid. *Ocean Modelling*, 143, 101466. <https://doi.org/10.1016/j.ocemod.2019.101466>
- Ye, F., Zhang, Y. J., Wang, H. V., Friedrichs, M. A. M., Irby, I. D., Altjeljevich, E., et al. (2018). A 3D unstructured-grid model for Chesapeake Bay: Importance of bathymetry. *Ocean Modelling*, 127, 16–39. <https://doi.org/10.1016/j.ocemod.2018.05.002>